Opportunities and challenges of Virtual Research Environments for biodiversity study

Paolo Colangelo, Bachir Balech, Angela Boggero, Pierluigi Bombi, Nicola Fiore, Alessandro Oggioni, Cataldo Pierri, Paolo Tagliolato

LifeWatch ITALIA

VREs in biodiversity domains

VREs have the potential to benefit research in all disciplines at all stages of research?

Web of Science
Showing 155 records for **TOPIC:** ("virtual research environment*")

Permeability of the biodiversity researchers community to VRE innovations

- The permeability level of ICT innovations is generally low biodiversity domain
- Few researchers are aware of the existence of VRE concept
- Many ecologists, zoologists, botanists are field researchers. Sometime they are intimidated also by "normal" computer and are frustated by using complex informatic tools
- Researchers want to use tools that has been proved to work (validated by publications)

Barriers to the use of VREs

- "…the majority of these systems (VREs) are not yet fully integrated into standard practices, tools and research protocols used by real life communities of practice."
- "This reluctance to migrate from traditional and consolidated research practices and facilities to the innovative ones promoted by VREs is among the most difficult barriers affecting the entire VRE domain."

Candela et al. (2013). Virtual Research Environments: An Overview and a Research Agenda. Data Science Journal. 12, pp.GRDI75–GRDI81

technology not reliable:	8
too difficult to use:	14
does not suit our research practice:	11
not enough technical support:	22
not enough institutional support for training etc.:	20
user community too small:	15
security/trust issues:	11
other:	36

Carusi & Reimer, 2010. . Virtual Research Environment Collaborative Landscape Study. JISC

LifeWatch VRE

Virtual Labs

Resources & Services / Catalogue of Virtual Labs

Alien and Invasive Species VRE

The LifeWatch Alien Species Virtual Research Environment (AS-VRE) has been built and equipped in order to developing systems that support the scientist's work for experimental researches on alien species arrival and spread into different types of ecosystems (aquatic and terrestrial). The AS-VRE is an example of the types of scientific studies that researchers on biodiversity and AS could undertake.

Would you like to know more about the Alien Species VRE, its services, get in contact with its coordination team, access the training resources, view the related publications and the showcases? Visit our dedicated minisite.

AS-VRE minisite ← Back

Search

gg/mm/aaaa

Text

Select Category

Search →

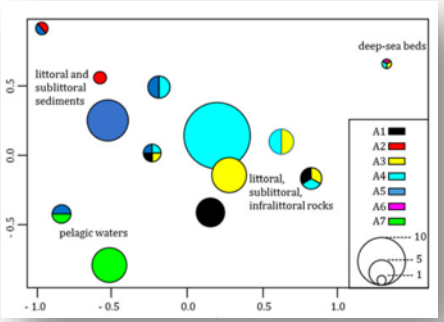
Category

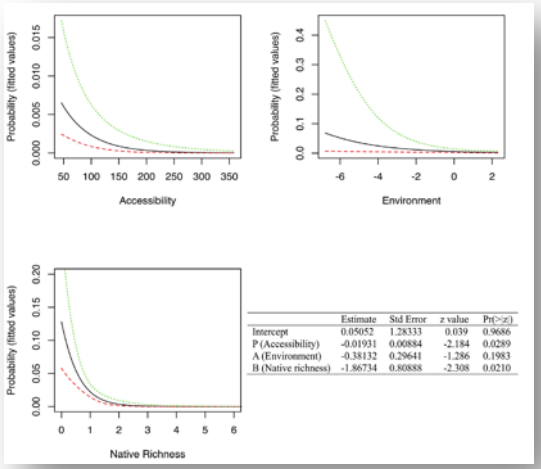
No Category

The Alien Species case study

AQUATIC CONSERVATION: MARINE AND FRESHWATER ECOSYSTEMS
Aquatic Conserv: Mar. Freshw. Ecosyst. **26**: 392–409 (2016)
Published online 20 February 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/aqc.2550

Ecosystem vulnerability to alien and invasive species: a case study on marine habitats along the Italian coast





REABIC
Aquatic Invasions (2017) Volume 12, Issue 3: 299–309
DOI: https://doi.org/10.3391/ai.2017.12.3.04

Open Access
Special Issue: Invasive Species in Inland Waters
Research Article

Alien species in Italian freshwater ecosystems: a macroecological assessment of invasion drivers

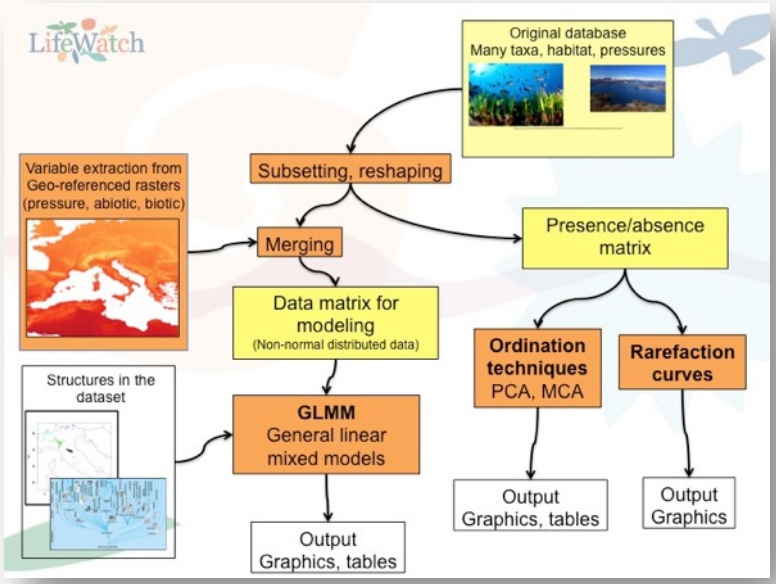
Ecological Indicators 91 (2018) 182–188
Contents lists available at ScienceDirect
Ecological Indicators
journal homepage: www.elsevier.com/locate/ecolind
ELSEVIER

Original Articles

Plant invasions in Italy: An integrative approach using the European LifeWatch infrastructure database

From statistical workflows to scientific workflows

- Allow users to replicate the workflow
- Allow user to modify the workflow
- Allow user to use their own data, enhancing modularity

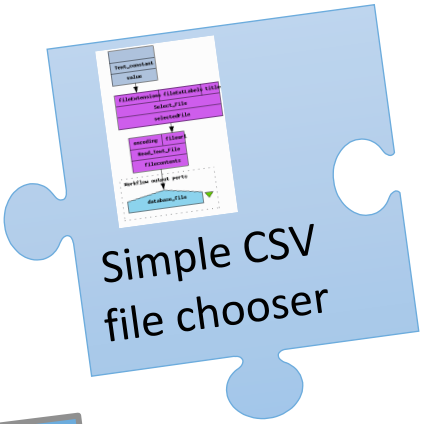


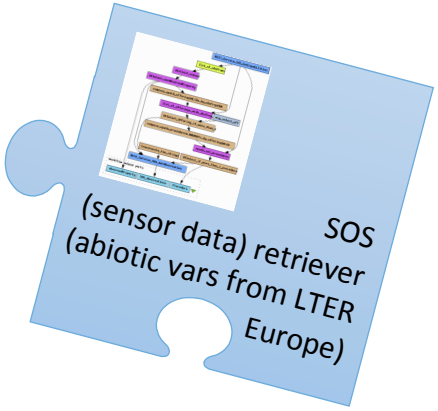

```
native_richness<-specnumber(native_table[sapply(native_table, class)!="factor"])
native_richness<-cbind(native_table[sapply(native_table,
class)=="factor"],native_richness)
##merge native+aliene
new_table<-merge(native_richness,alien_richness,all.x=T,all.y=T)
new_table[is.na(new_table)]<-0

# remove unnecessary files
rm("alien_richness")
rm("native_richness")
```

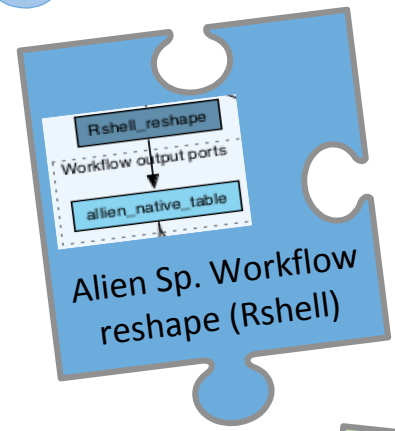
```
##############################################################################
##
#####  Step 2: Generalized Linear Mixed Model (GLMM) fitting usign the lme4 package
#####
##############################################################################
##


# now we are ready to fit our model. We will use a generalized linear mixed models in
order to take into account the structure of our new dataset. Taxonomic group and
locality are not the focus of our investigation but largely influence our sampling. We
will include these two factor in the random effect.

# First fit full model (a negative bionomial family is assumet for richness data)
gfit_Eu_Ri <- glmer.nb(alien_richness ~native_richness+ EunisL1 +(1|
family)+(1|locality), data= new_table)

# automatically calculate best model according to AIC
library(MuMIn)
options(na.action = "na.fail")
ms1<-dredge(gfit_Eu_Ri)
ms1; # the full model has the highet AICc support

# fit the best model accordin to AIC
mod.fit<-glmer.nb(as.formula(getCall(ms1,1)), data = new_table)

# results
summary(mod.fit) #table.
visreg(mod.fit,trans=exp,nn=101,alpha=1,rug=F,partial=T) #graph
```

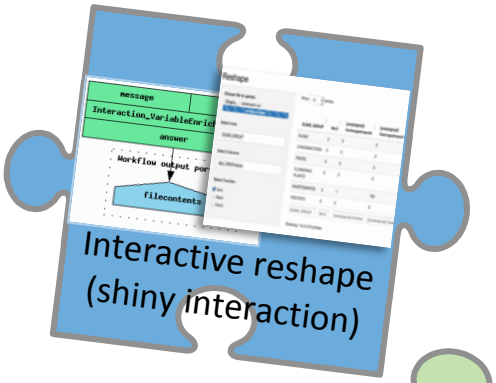
```
####################################
##### Step 3: RAREFACTION CURVES #####
####################################

library(vegan)

asfw<-droplevels(subset(freshwater, alien=="1"))
tabella_aspecie<-table(asfw $EunisL1, asfw $scientificname)
rarecurve(tabella_aspecie)
```

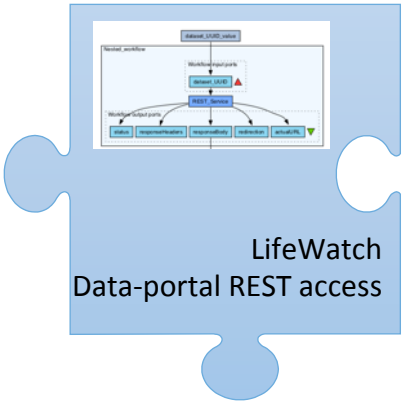
Dividi et impera!

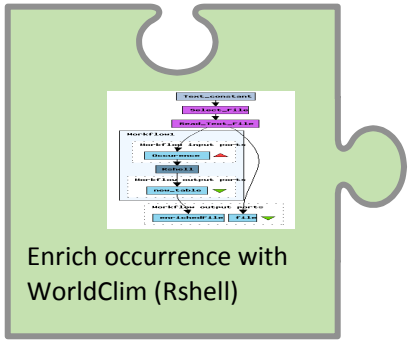


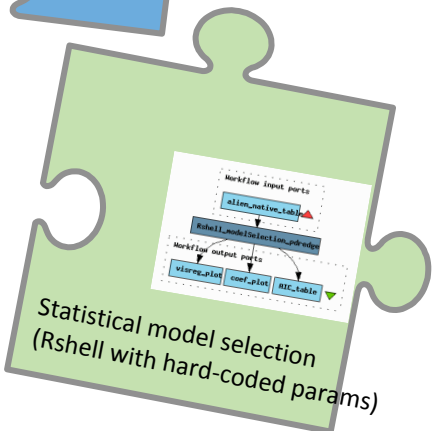


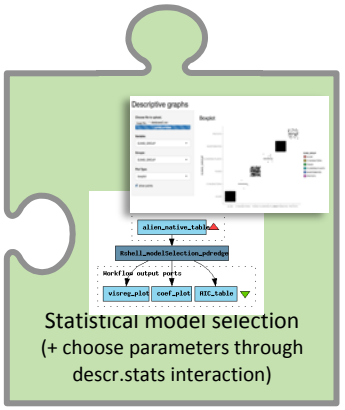






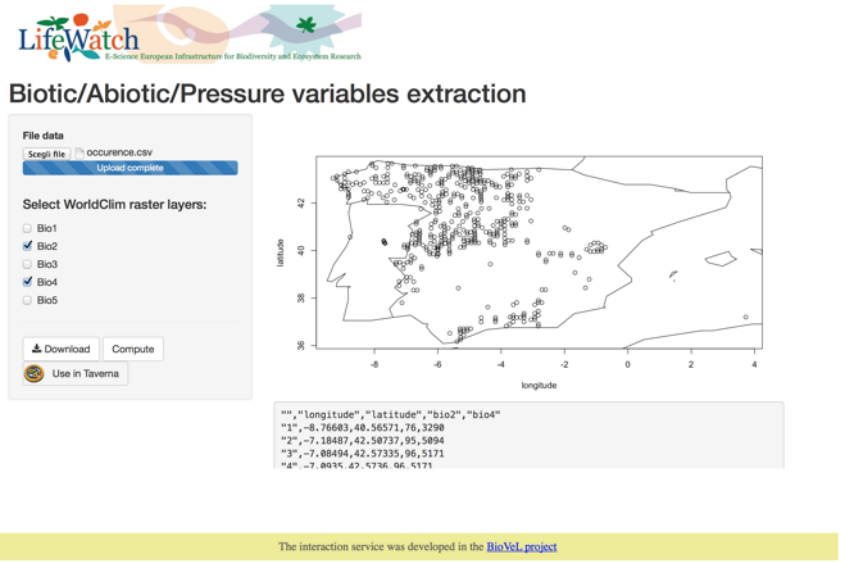






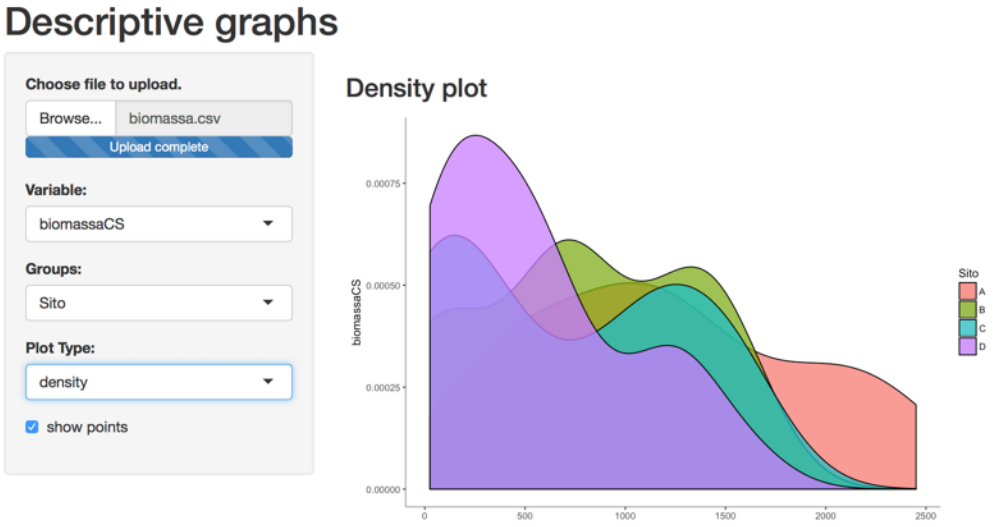
Enhancing the agent-user interaction

- Progresses: reusable web-application interaction services with **advanced interactivity** and **processing**
- **Taverna aware** *Shiny (R) applications*

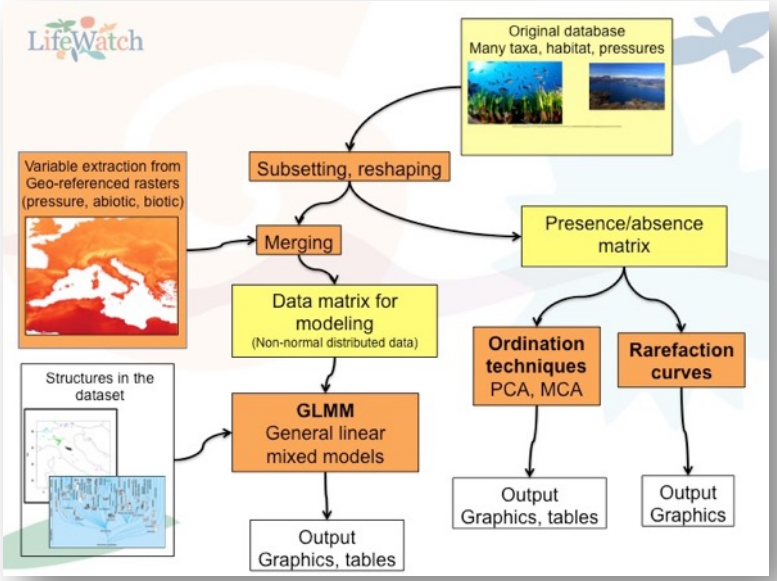


Some example:

- Enrich species occurrence with abiotic data from WorldClim
- Choices of parameter through *descriptive statistics*
- Dynamically reshape data



From R to Taverna

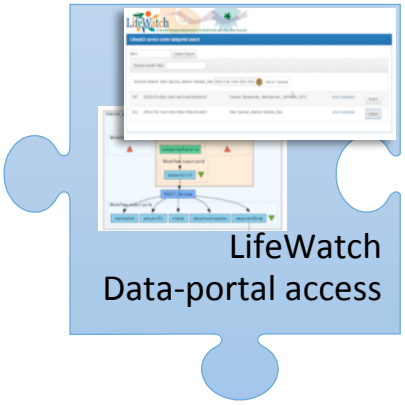



```
library(shiny)
library(maps)
library(raster)

shinyServer(function(input, output, session) {

  Data <- reactive({
inFile<-input$file
    if (is.null(inFile)) {
      return(NULL)
    } else {
      d<-read.csv(inFile$datapath)
d
    }
  })



  observe({
    data <-Data()
    updateSelectInput(session, "x", choices = names(data))
    updateSelectInput(session, "y", choices = names(data))
  })
```

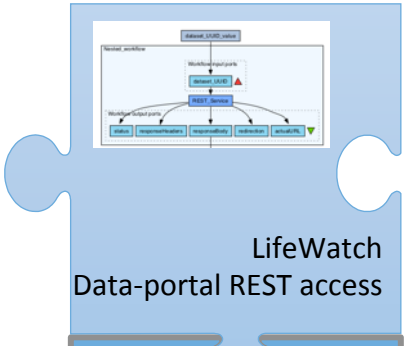
Taverna service oriented scientific workflow

Implement R scripts via Rshell and Rserve.
Connect R scripts with other services

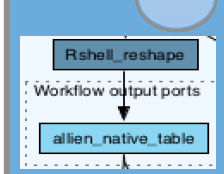
Re-combination



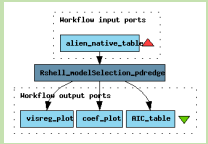
LifeWatch
Data-portal access



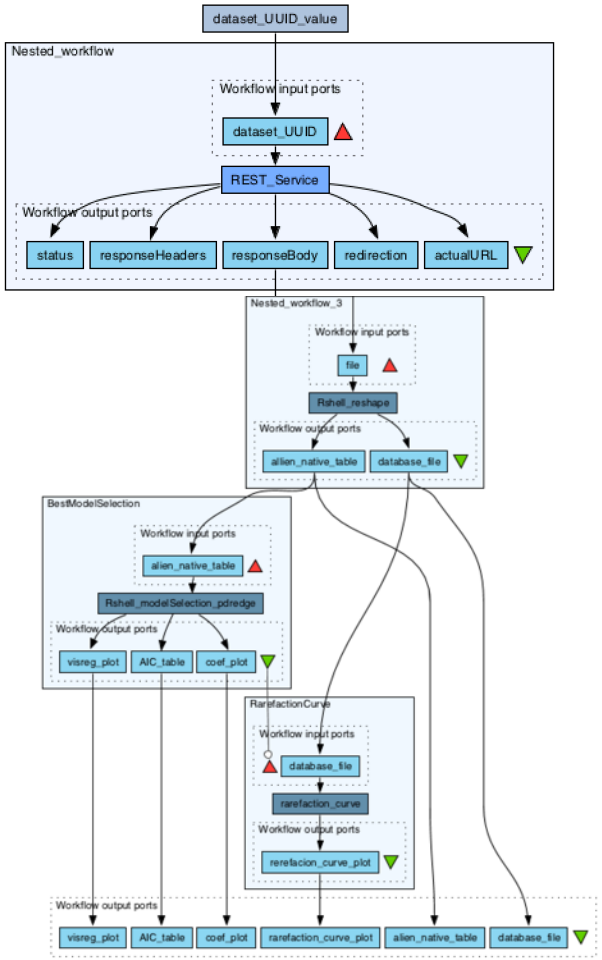
LifeWatch
Data-portal REST access



Alien Sp. Workflow reshape (Rshell)



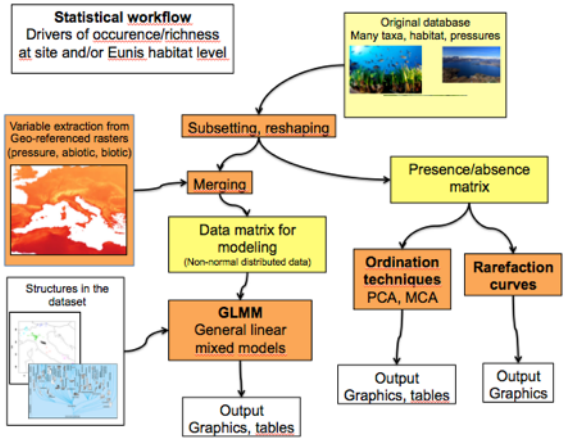
Statistical model selection (Rshell with hard-coded params)

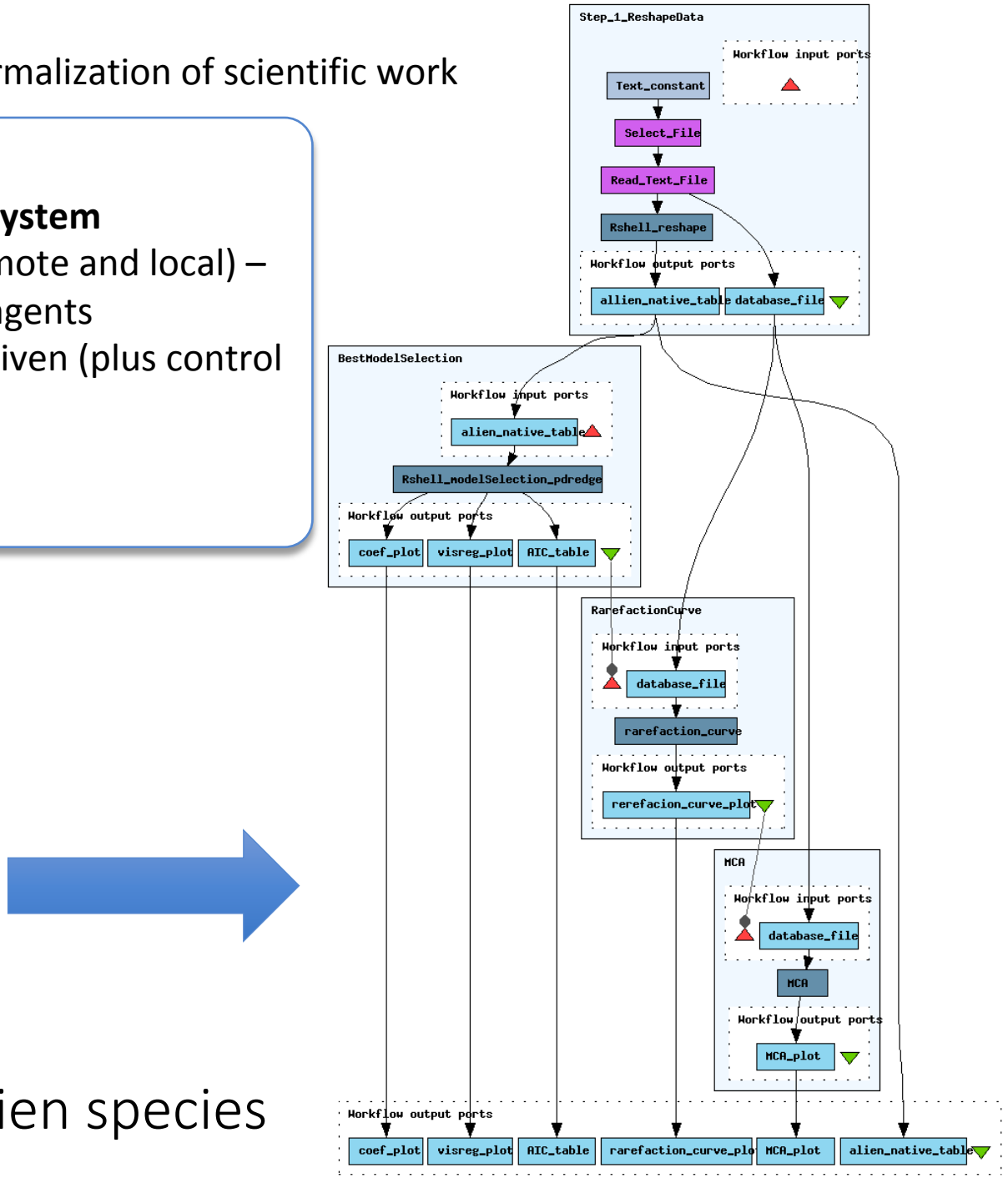


The workflow as a runnable formalization of scientific work

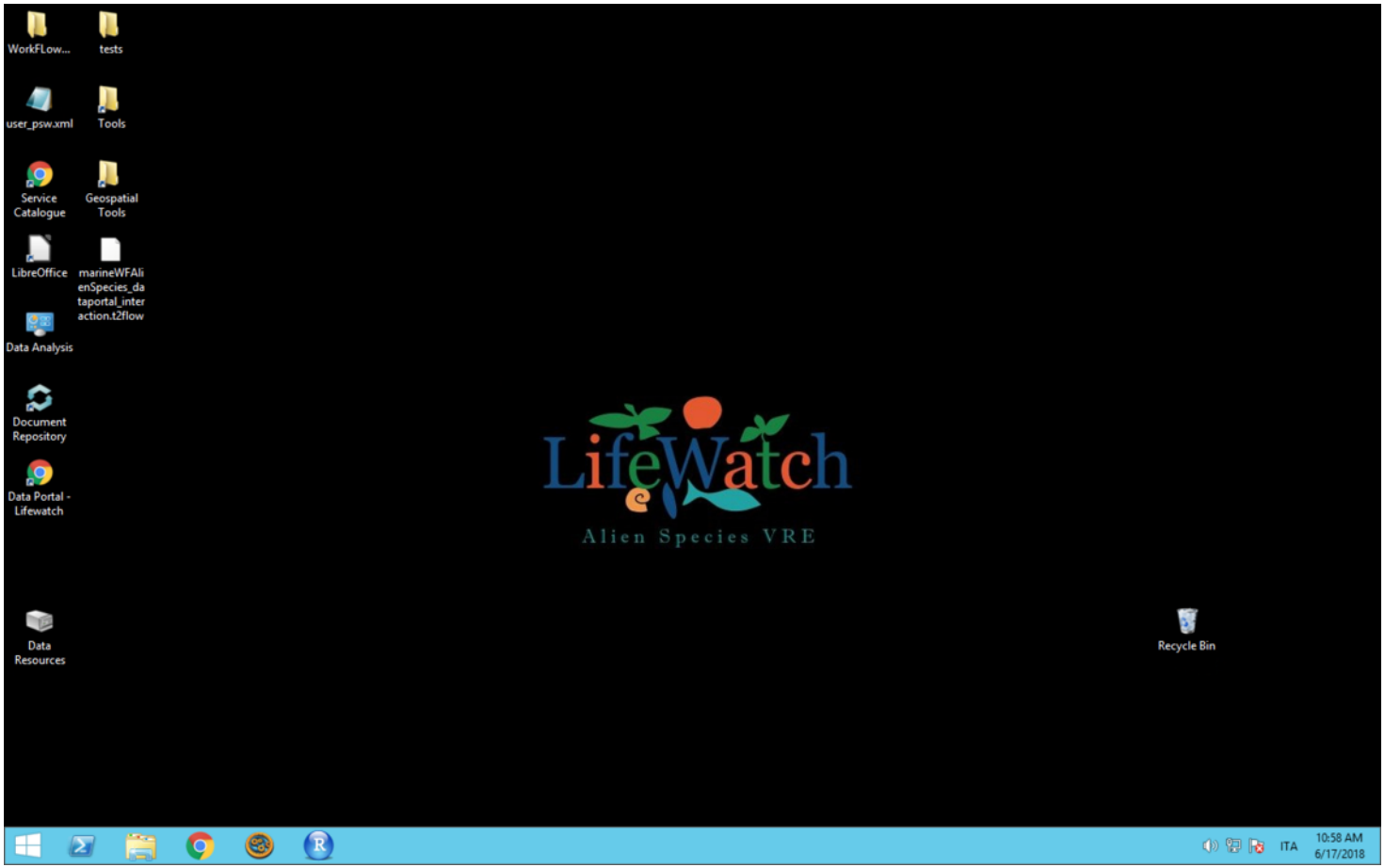
Taverna workflow system

- Composed by “services” (remote and local) – configurable assignment to agents
- Workflow execution: data driven (plus control links)
- Nested workflows





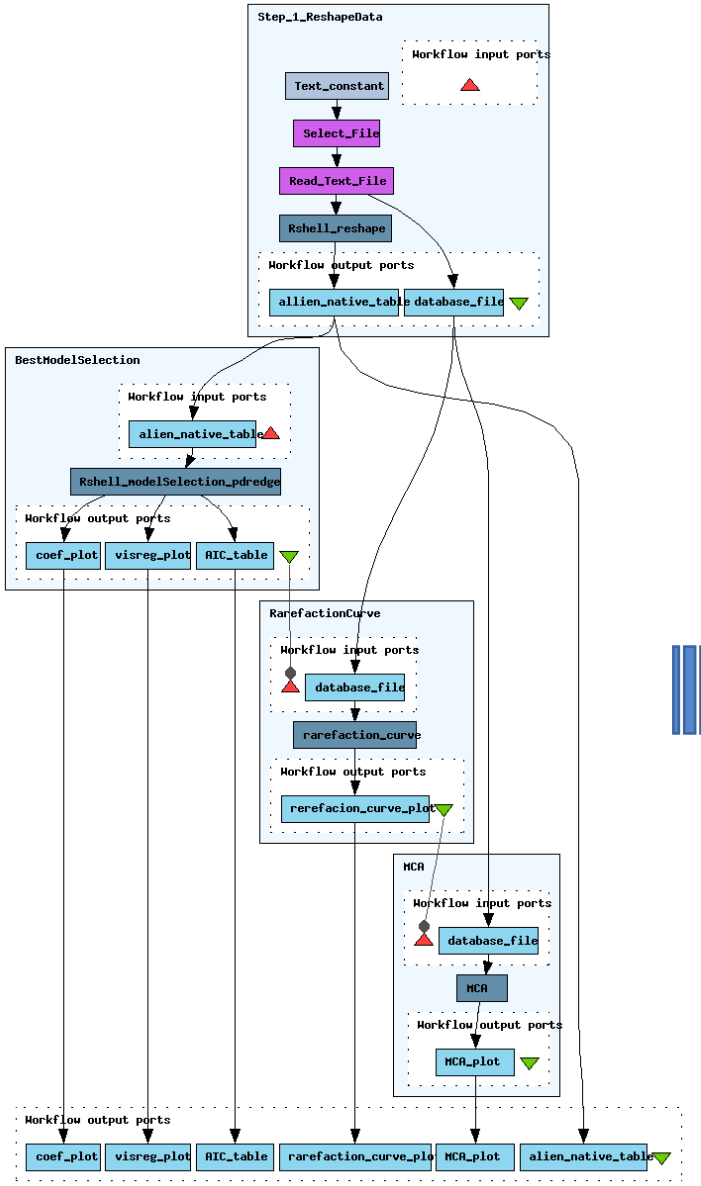
Example workflow: alien species



- Possibility to replicate the AS showcase analysis
- Possibility to use new data that are stored in the LW dataportal
- Allow some degree of workflow manipulation

- Needs to develop new app or web services
- Authentication with multiple users on the same project is triky
- Taverna is not for all

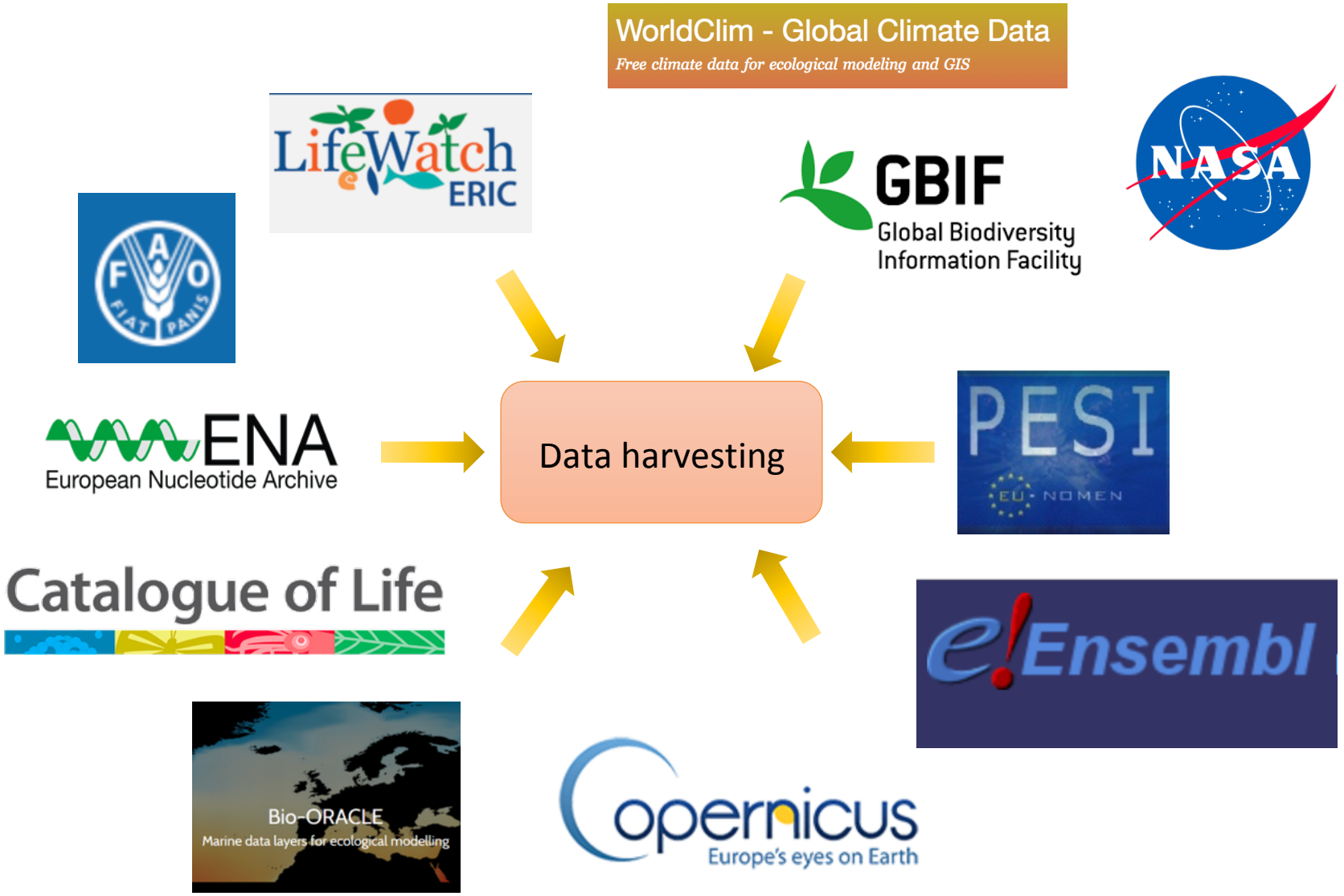
Complex workflows and researchers…

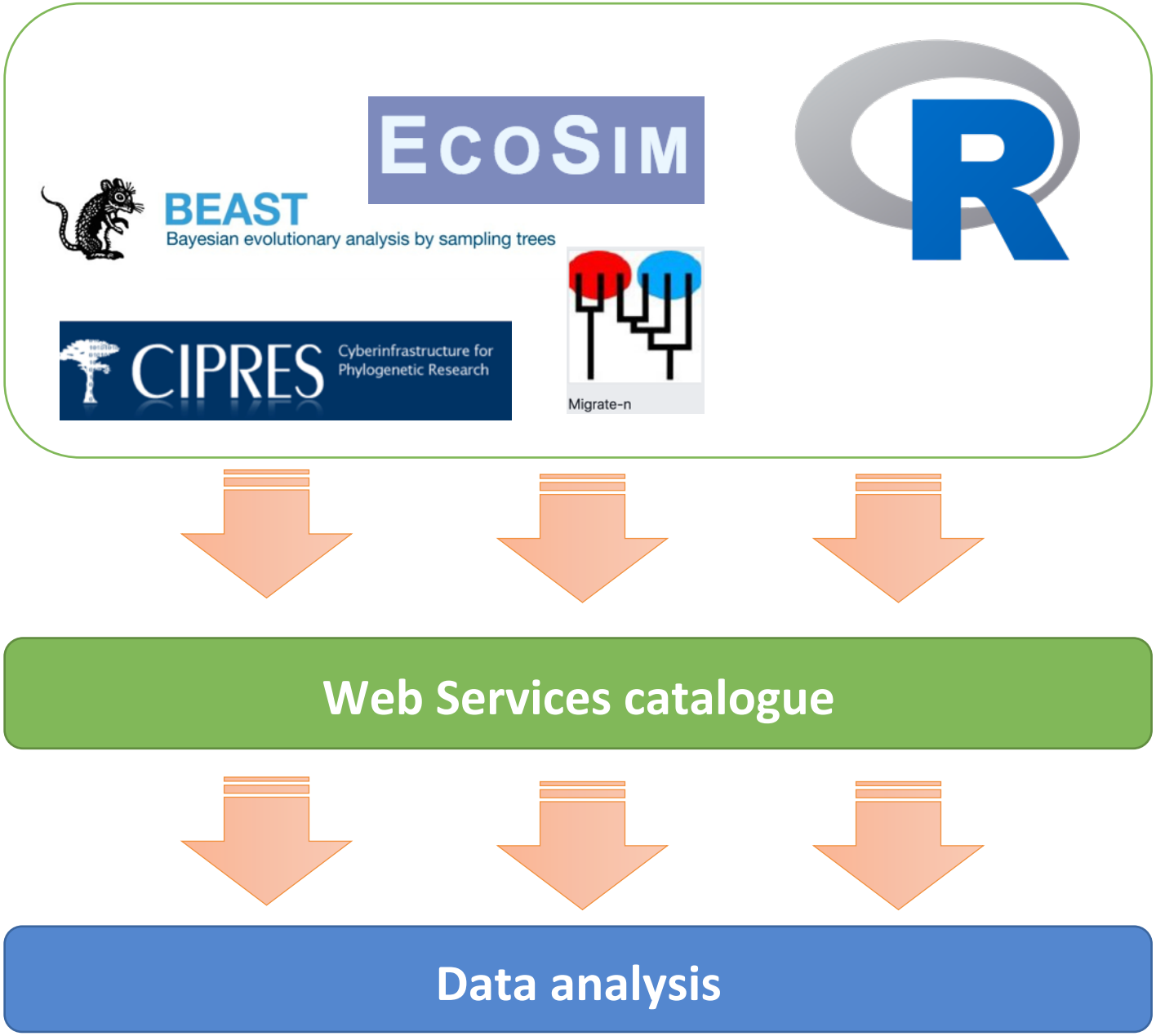


What biodiversity researchers want from a VRE?

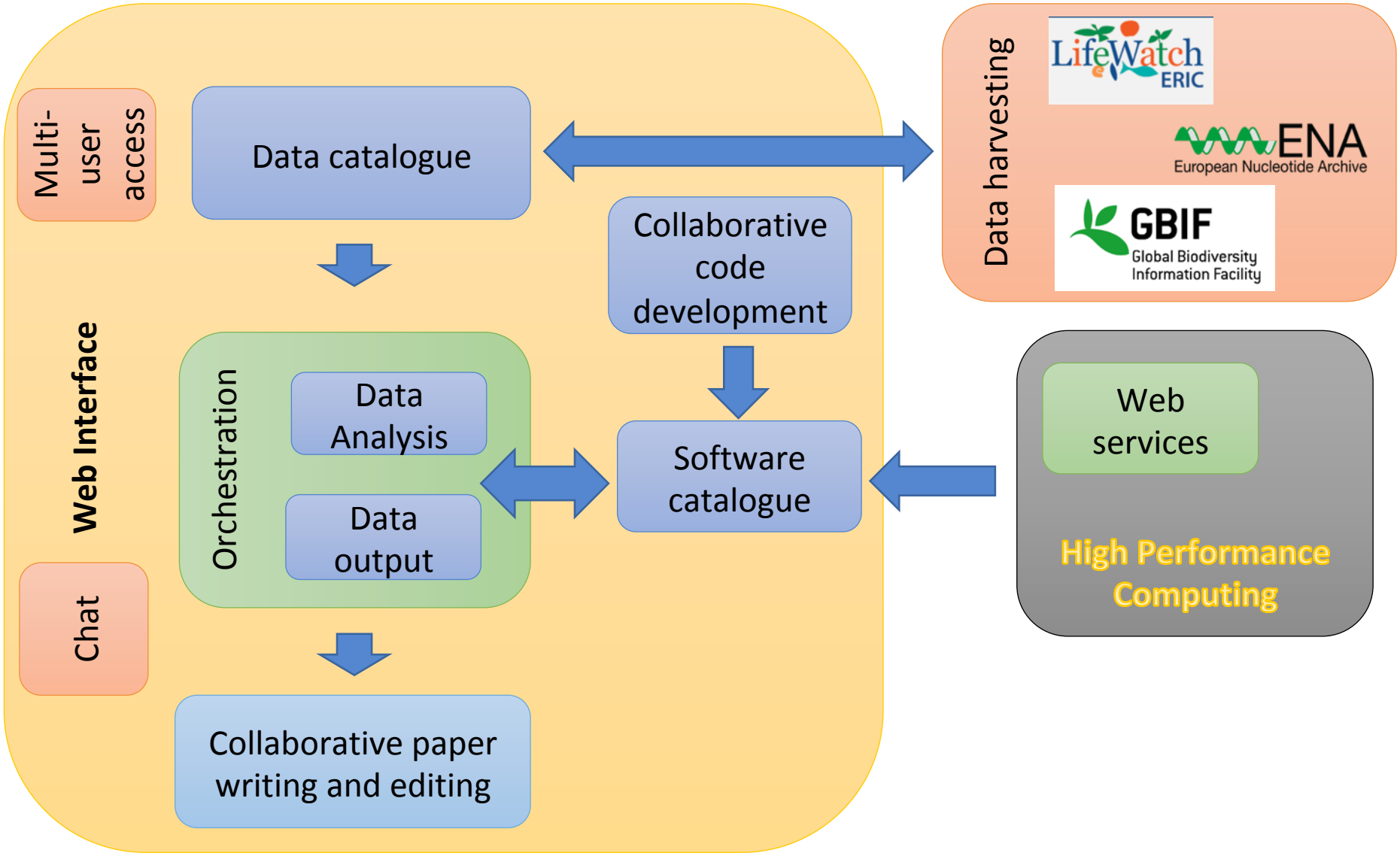
1. Access to a wide range of biodiversity data (data catalogues)
2. High computing capacity (HPC clusters);
3. Access to a wide range of software validated by the scientific community;
4. A collaborative environment
5. User friendly interface, accessible via web technologies

Access to biodiversity data




EcoSim
BEAST
Bayesian evolutionary analysis by sampling trees
R
CIPRES
Cyberinfrastructure for Phylogenetic Research
Migrate-n
Web Services catalogue
Data analysis

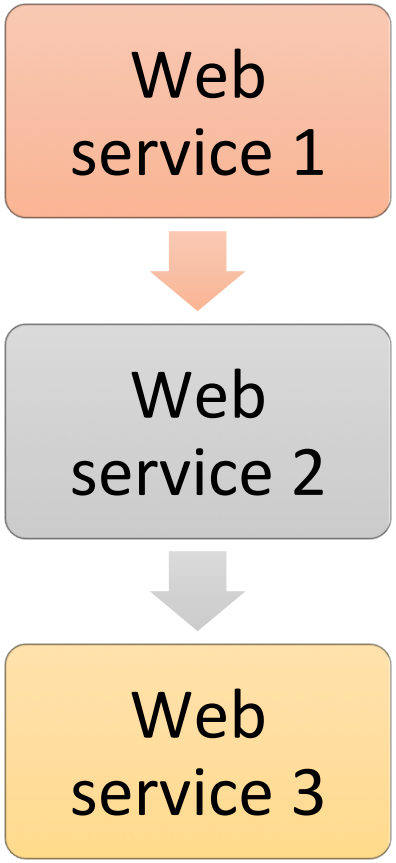
VRE-LW architecture



Orchestrating web services

Taverna

Galaxy PROJECT



Different solutions possible, but be aware always to the researchers sensibility…

Conclusions

- The access to VREs should lead to faster research results by providing data, tools, computational resources and a collaborative environment
- If the intention is to improve the research process and not simply to show technologies for their own sake, the requirements of users (researchers) must drive the VRE developments

Conclusions

If the intention is to improve the research process and not simply to show technologies for their own sake, the requirements of users (researchers) must drive the VRE developments

Virtual Research Environments should be designed, since the beginning, to *promote uptake, ensure usability, and guarantee sustainability*. These three aspects form a virtuous circle that, if properly managed, ensure the success of a specific VRE.

As regards *usability*, **Virtual Research Environments building should be mainly a community building process rather than a technology development process**. This implies that the focus should be primarily on using technology to identify and rationalise workflows, procedures,and processes characterising a certain research scenario rather than having technology invading the research scenario and distracting effort from its real needs.

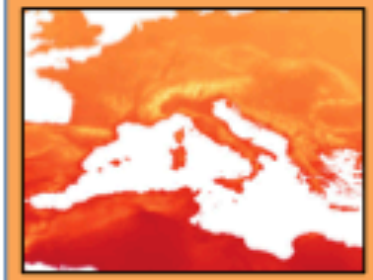
Candela et al. (2013). Virtual Research Environments: An Overview and a Research Agenda. Data Science Journal. 12, pp.GRDI75–GRDI81

Statistical workflow
Drivers of occurence/richness at site and/or Eunis habitat level

Original database
Many taxa, habitat, pressures

Subsetting, reshaping

Variable extraction from Geo-referenced rasters (pressure, abiotic, biotic)



Merging

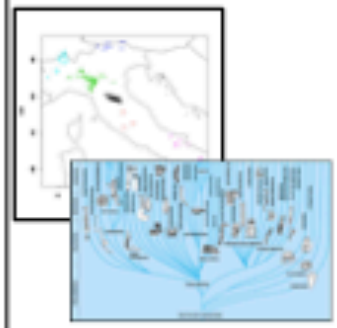
Presence/absence matrix

Data matrix for modeling
(Non-normal distributed data)

Ordination techniques
PCA, MCA

Rarefaction curves

Structures in the dataset



GLMM
General linear mixed models

Output
Graphics, tables

Output
Graphics, tables

Output
Graphics